

AK035300 3224 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male urinary bladder cDNA, RIKEN full-length
enriched library, clone:953001P19 product:similar to CASPASE
RECRUITMENT DOMAIN PROTEIN 12 (ICE-PROTEASE ACTIVATING FACTOR
(IPAF) (CARD, LRR, AND NACHT-CONTAINING PROTEIN) (CIAN PROTEIN)
[Homo sapiens], full insert sequence.
AK035300
AK035300.1 GI:26084515
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
PUBMED 10349636

REFERENCE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED 11042159

REFERENCE

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED 11076861

REFERENCE

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuura, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerle, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wyehaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
MEDLINE
PUBMED 11217851

REFERENCE

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
11217851
MEDLINE
PUBMED 12035660

6 (bases 1 to 3224)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saich, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

Direct Submission

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

FEATURES

source

Location/Qualifiers
1 .3224
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:9530011P19"
/db_xref="taxon:10090"
/clone="9530011P19"
/sex="male"
/tissue type="urinary bladder"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
<1 .2693
/note="putative"

misc_feature

similar to CASPASE RECRUITMENT DOMAIN PROTEIN 12 (ICE-PROTEASE ACTIVATING FACTOR) (IPAF) (CARD, LRR, AND NACHT-CONTAINING PROTEIN) (CLAN PROTEIN) (Homo sapiens) (SWISSPROT|Q9NPP4, evidence: FASTA, 76%ID, 87.5%length, match=2689)"
3205..3210
polyA_signal
3224
polyA_site
/note="putative"

BASE COUNT 929 a 754 c 779 g 762 t

ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 3224
Score: 3496.00 Matches: 681
Percent Similarity: 85.06% Conservative: 82
Best Local Similarity: 75.92% Mismatches: 133
Query Match: 66.25% Indels: 1
DB: 11 Gaps: 0

US-09-864-921-97 (1-1024) x AK035300 (1-3224)

QY 129 AsnLeuYsSerThrPheThrGluProIleLeuTrpArGlyAspGlnHisHisArg 148

Db 2 AATCTGGAGAAACCTTCACAGAACCTATCATCTGGAGAGAGGACCATCGTCATCCCGT 61

QY 149 ValGluGlnLeuThrLeuAsnGlyLeuGlnAlaLeuGlnSerProCysIleIleGlu 168

Db 62 GTGGAGCAGCTGACTTTTGGGCGAGCCTGTCTGAGGCTCTGAAGAGCCCTCGCTGATTGAA 121

QY 169 GlyGluSerGlyLyGlyLySerThrLeuLeuGlnArgIleAlaMetLeuTrpGlySer 188

Db 122 GCGGAGTCTGGCAAGGGAAGTCCACCTGCTGCAGAGAAATCGCTATGCTCTGGGCGCTCT 181

QY 189 GlyLysCysLysAlaLeuThrLysPheLysPheValPhePheLeuArgLeuSerArgAla 208

Db 182 GGGGGCTCGAGGGCTCTGAAGGGGTTCAGATTAGTCTTCTTCATCCACCTGAGAGAGCGCC 241

QY 209 GlnGlyGlyLeuPheGluThrLeuGlnLeuGlnLeuLeuAspIleProGlyThrIleArg 228

Db 242 ATGGGGGAGCTATTTCGAACAACACTGTACGATCAGCTCTCTGAACATACACCGCTTCATCAGC 301

QY 229 LysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAsp 248

Db 302 AAGCCGACCTTCAAGGCTCTGCTGCTGAAGCTACACAAGAGGCTCCTCTTCTTCTCGAT 361
Qy 249 GlyTyrAsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysGluLeu 268
Db 362 GGTTCATGAATTCATCCCAAGAACTCCCAAGAAATGAAGCCCTGATAAGGAAAC 421
Qy 269 HisArgPheLysAsnMetValIleValThrThrThrGluCysLeuArgHisIleArg 288
Db 422 CATCGCTTCAAGAACATGCTGTCATTCACCAACCACCGAGTCTGAGGCATATCAGA 481
Qy 289 GlnPheGlyAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAlaLeu 308
Db 482 CATGTGGCGCCCTGACTCGGAGGTGGAGATATGACCAAGACAGAGTCCCAAGATCTC 541
Qy 309 IleArgGluValLeuIleLysGluLeuAlaGluGlyLeuLeuGlnIleGlnLysSer 328
Db 542 ATCGAGCGAGTCTGCTGCTGATCAGGTTGAACGCTGTGGGCCCAAAATCCAGGAGTCC 601
Qy 329 ArgCysLeuArgAsnLeuMetLysThrProLeuPheValIleThrCysAlaIleGln 348
Db 602 AGGTGCTGAGAAATCTGATGAAGACCCCTCTCTTCGTGCTGATCACCTGTGCAATTCAG 661
Qy 349 MetGlyLysSerGluPheHisThrHisThrGlnThrThrLeuPheHisThrPheTyrAsp 368
Db 662 ATGGGCAGACAGGAATTCAGAGCTCACACCAACCATGCTGTCCAAACCTTCTACGAC 721
Qy 369 LeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaAlaSerAspPheIleArg 388
Db 722 CTCCTGATACAGAAACAGCCACAGATATAGAGTGGAGCTTCAGGTGATTTGCCAGG 781
Qy 389 SerLeuAspHisArgGlyAspLeuAlaLeuGluGlyValPheSerHisLysPheAspPhe 408
Db 782 AGCTTAGACTACTGTGGAGCTTGGCCCTAGAGGTGTGTTGCCCAACAATTCATTTT 841
Qy 409 GluLeuGlnAspValSerValAsnGluAspValLeuLeuThrThrGlyLeuLeuCys 428
Db 842 GAACCCGAGCATGGGTCCAGCATGAACGAGGACGCTCTGGTGACAATAGGGCTCCTCTGT 901
Qy 429 LysTyrThrAlaGlnArgPheLysProLysTyrLysPhePheHisLysSerPheGlnGlu 448
Db 902 AAGTACACAGCTCAGAGGCTGAAGCCACGTTAAATCTTTCAATAATCATTTTCAGGAG 961
Qy 449 TyrThrAlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThr 468
Db 962 TACACGGCAGGTGGAGACTCAGCAGTTGCTGACGCTCAAGAGCCAGAGGAGGTGACC 1021
Qy 469 LysGlyAsnGlyTyrIleuGlnLysMetValSerIleSerAspIleThrThrTyrSer 488
Db 1022 AAAGGGACACAGCTACTTAAACAAAATGGTTTCCATCTCTGACATCACATCCCTATATGC 1081
Qy 489 SerLeuLeuArgTyrThrCysGlySerSerValGluAlaThrArgAlaValMetLysHis 508
Db 1082 AATCTGCTCTCTACACGTTGGGTGCTCCAGAAAGCAACAGGCGGCTCATAGGCAC 1141
Qy 509 LeuAlaAlaValTyrGlnHisGlyCysLeuLeuGluGlyLeuSerIleAlaLysArgProLeu 528
Db 1142 CTTCGATGTTTATCAGACCGCAGCCTCAGAGACTTTCAGTACCAAGAGGCTCTC 1201
Qy 529 TrpArgGlnGluSerLeuGlnSerValLysAsnThrThrGluGlnGluIleLeuLysAla 548
Db 1202 TGGAGGACAGGAATCAATCCAGAGTCTGAGAAATACCATGTGAGCAAGATGTTCTGAAAGCC 1261
Qy 549 IleAsnIleAsnSerPheValGluCysGlyIleHisLeuTyrGlnGluSerThrSerLys 568
Db 1262 ATCAATGTAAATTCCTCTGATAGAGTGGCATCAATTTGTTCTCAGAGAGTATGTCTAAA 1321
Qy 569 SerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlyLysSerLeuTyrIleAsnSer 588
Db 1322 TCAGACCTGAGCCAGAAATTCAGAGCTTTCTTTCAAGGTAAAGTTTATACATCAACTCA 1381
Qy 589 GlyAsnIleProAspTyrLeuPheAspPhePheGluHisLeuProAsnCysAlaSerAla 608
Db 1382 GAGAACATCCCTGACTATTATTATTTGACTTCTTTTGAATACTTGCCTAAATTTGTCAAGGCCA 1441

Qy 609 LeuAspPheIleLysLeuAspPheTyrGlyGlyAlaMetAlaSerTrpGluLysAlaAla 628
Db 1442 TTGGACTTCGTGAAGTTGGATTTCTATGAAGAGCTACAGAGTACAGAGCAAGCAGNA 1501
Qy 629 GluAspThrGlyGlyIleHisMetGluAlaProGluThrTyrTrpIleProSerArgAla 648
Db 1502 GAGAATGTCCTGGAGTTTACACAGAGGGCCCTCAGAAACCTACATATCCCCCGGGCT 1561
Qy 649 ValSerLeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValThrLeuArgAsp 668
Db 1562 GTGTCTTTGTTCTTCACTGGAAGCAGGAATTCAGAGCTCTAGAGGTACACTTCGAGAT 1621
Qy 669 PheSerLysLeuAsnLysGlnAspIleArgTyrIleGlyLysIlePheSerSerAlaThr 688
Db 1622 ATTACAGTTGTAATAGCAAGATATCAATATCTGGGGAAGATATTAGCTCTGCCACC 1681
Qy 689 SerLeuArgLeuGlnIleLysArgCysAlaGlyValAlaGlySerLeuSerLeuValLeu 708
Db 1682 AACCTCGGCTGCATATCAAGAGATGTGCAGCATGTGGTGGAAAGACTCAGCTCAGTCTC 1741
Qy 709 SerThrCysLysAsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAsp 728
Db 1742 AGAACCCTCAAGAACATCATACCTCATGTGGTGAAGCCAGTCCCTCACCACGGATGAC 1801
Qy 729 GluArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsn 748
Db 1802 GAACAGTACATCATCATCTGTGACAGCCCTCAGAACTTAAGTATTTCACCGTTGCACACT 1861
Qy 749 GlnArgLeuProGlyLysLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLysLeu 768
Db 1862 CAACAGCTGCCAGGTGGTCTGATTGACAGCTTGGTATCTCGAAGAACTCGAGAGACTC 1921
Qy 769 IleMetAspAsnIleLysMetAsnGluAspAlaIleLysLeuAlaGluGlyLeuLys 788
Db 1922 ATACTGGATGATCAGGATGAACGAGGAAGATGCTAAACAACTAGCCGAGGCTTACGA 1981
Qy 789 AsnLeuLysLysMetCysLeuPheHisLeuThrHisLeuSerAspIleGlyGluGlyMet 808
Db 1982 AGCCTGAAGAGATGCGTTTACTCATTTGACTCATTTGTCTGACATTTGGGGGGGGATG 2041
Qy 809 AspTyrIleValLysSerLeuSerGluProCysAspLeuGluGluIleGlnLeuVal 828
Db 2042 GACTACATAGTCAAGTCTCTCTCAGAAAGATCTCTGATCTCCAAGAGATGAAGTTGGTG 2101
Qy 829 SerCysCysLeuSerAlaAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuVal 848
Db 2102 GCCTGCTCTGACTGCAAACTCTGTGAAAGTTCTAGCACAGAACTCTTCACAATTTGATC 2161
Qy 849 LysLeuSerIleLeuAspLeuSerGluAsnTyrIleuGluLysAspGlyAsnGluAlaLeu 868
Db 2162 AAGCTGACATCTTTGATATATCAGAAATTTACCTGGAAAGAGTGGGAATGAAGCTCTA 2221
Qy 869 HisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAlaLeuMetLeuProTrp 888
Db 2222 CAGGAACATGATCGCAGGCTTGGCGTCTCGGAGAGCTCACTACATTTGATGCTGCTGG 2281
Qy 889 GlyCysAspValGlnGlySerLeuSerSerLeuLysHisLeuGluGluValProGln 908
Db 2282 TGCTGGGATGTGCACACCCAGCTGCCAAGCTGTTGAAGCAGATTGGAGGGGCCCCAGGA 2341
Qy 909 LeuValLysLeuGlyLeuLysAsnTrpArgLeuThrAspThrGluIleArgIleLeuGly 928
Db 2342 CTTCGCAAACTTGGATTGAAAACCTGGAGACTCAGAGACGAAGATTAAGATTAGGT 2401
Qy 929 AlaPhePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArg 948
Db 2402 GAATTTCTGGAGATCAATCTCTGAGACCTTGCAGCAGTTGGATTTAGCGGGGCACTGT 2461
Qy 949 ValSerSerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLeuVal 968
Db 2462 GTGAGCAGTACGGATGCTTTTACTTTCATGAATGTGTTTGAATCTCTGAAGCAGTTAGTG 2521

```

QY 969 PhePheAspPheSerThrLysGluPheLeuProAspProAla-LeuValArgLysLeuSe 988
Db 2522 TTTTITGACTTTAGCAGCTGAGAGTCTTACCGATGAGCCACTGGTGAGGAAACTTAG 2581
QY 988 rGlnValLeuSerLysLeuThrPheLeuGlnGluAlaArgLeuValGlyTrpGlnPheAs 1008
Db 2582 TCAAGTGTATCAAGTTAACTCTTCTGCAAGAGGTAAAGCTCACGGGCTGGAGTTTGA 2641
QY 1008 pAspAspLeuSerValIleThrGlyAlaPheLysLeuValThrAla 1024
Db 2642 TGACTATGATATTAGCGCTATTAAAGGCACCTTTAAACTAGTACTGCT 2690

RESULT 2
AQ309404/c AQ309404 552 bp DNA linear GSS 22-DEC-1998
LOCUS CITBI-E1-2528J13.TF CITBI-E1 Homo sapiens genomic clone 2528J13,
DEFINITION genomic survey sequence.
ACCESSION AQ309404
VERSION AQ309404.1 GI:4041438
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
JOURNAL Map Building
COMMENT Unpublished
Other GSSs: CITBI-E1-2528J13.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbs@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tcdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES             location/Qualifiers
source               1..552
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2528J13"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
BASE COUNT          144 a 121 c 130 g 157 t
ORIGIN

Alignment Scores:
Pred. No.:          2,25e-96          Length:          552
Score:              952.00            Matches:         183
Percent Similarity: 100.00%           Conservative:    0
Best Local Similarity: 100.00%         Mismatches:      0
Query Match:        18.04%            Indels:          0
DB:                  28                Gaps:            0

US-09-864-921-97 (1-1024) x AQ309404 (1-552)

QY 477 MetValSerIleSerAspIleThrSerThrTyrrSerSerLeuLeuArgTyrrThrCysGly 496
Db 550 ATGGTTTCCATTTCGACATTACATCCACTTAGCAGCTCTCGGTACACCTGTGGG 491
QY 497 SerSerValGluAlaThrArgAlaValMetLysHisLeuAlaValTyrrGlnHisGly 516

```

```

Db 490 TCACTCTGGAGCCACCAGGCGCTGTATGAAGCACCTCGCAGCAGTGTATCAACACGGC 431
QY 517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnGlnSerLeuGlnSer 536
Db 430 TGCCTTCTCGGACTTTCCATCGCCCAAGAGGCGCTCTCTGGAGACAGGAATCTTTGCAAAAGT 371
QY 537 ValLysSerThrThrGluGlnGluIleLeuLysAlaIleAenIleAenSerPheValGlu 556
Db 370 GTGAAAACACCACTGAGCAAGAAATCTGAAGCCATAAATCAATTCCTTTGTAGAG 311
QY 557 CysGlyIleHisLeuTyrrGlnGluSerThrSerLysSerAlaLeuSerGlnGluPheGlu 576
Db 310 TGTGGCATCCATTATATCAAGAGAGTACATCCAAATCAGCCCTGAGCCAGAAATTTGAA 251
QY 577 AlaPhePheGlnGlyLysSerLeuTyrrIleAenSerGlyAenIleProAspTyrrLeuPhe 596
Db 250 GCTTTCTTTCAAGGTAAAGCTTATATATCACTCAACTCAGGGAACATCCCCGATTACTTTT 191
QY 597 AspPhePheGluHisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPhe 616
Db 190 GACTTCTTTGAACATTTGCCCAATTTGTCAAGTGTCTGTGACTTCAATTAACCTGGACTTT 131
QY 617 TyrGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrGlyGlyIleHisMet 636
Db 130 TATGGGGAGGATATGGCTTCATGGGAAAGGCTGCAGAGACACAGGTGGATCCACATG 71
QY 637 GluGluAlaProGluThrTyrrIleProSerArgAlaValSerLeuPhePheAenTrpLys 656
Db 70 GAAGAGGCCCCAGAAACCTACATTTCCACAGCAGGCGTGTATCTTTGTTCTTCAACTGGAAG 11
QY 657 GlnGluPhe 659
Db 10 CAGGAATTC 2

RESULT 3
BH293386
LOCUS BH293386
DEFINITION CH230-44G15.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION BH293386
VERSION BH293386.1 GI:17205794
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsagaye,G., Geer,K., Shvartsbeyn
,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Rigge,F., de
Jong,P. and Frazer,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL Unpublished
COMMENT Other GSSs: CH230-44G15.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.html). BAC end
page: http://www.tigr.org/tcdb/bac_ends/rat/bac_end_intro.html
Plate: 44 row: G column: 15
Seq primer: T7
Class: BAC ends.

FEATURES             location/Qualifiers
source               1..697

```

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-44G15"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
Pieter de Jong"
BASE COUNT      182 a   165 c   190 g   160 t
ORIGIN

Alignment Scores:
Pred. No.:      3,51e-94      Length:      697
Score:          934.00      Matches:      184
Percent Similarity: 87.77%      Conservative: 17
Best Local Similarity: 80.35%      Mismatches: 28
Query Match:      17.70%      Indels:      1
DB:              28          Gaps:      0

US-09-864-921-97 (1-1024) x BH293386 (1-697)

QY      266  LysGluAsnHisArgPheLysAsnMetValIleValThrThrThrGluCysLeuArg 285
DB      11  AAGGAAACCATCG-TTTAAGAACATGGTCAATGTCCACCACCGAGTGCCTGAGG 69

QY      286  HisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAla 305
DB      70  CACATCAGACACGTTGGCGCCCTGACTGTGGAGTGGGAGATGATGACCAAGACACGCGC 129

QY      306  GinAlaLeuIleArgGluValIleLysGluLeuAlaGluClyLeuLeuGlnIle 325
DB      130  CGAGTTCTCATCGGGAAGTGTGATAATGAACCTGGCTGAAGGCTGTGTGTCCAGATG 189

QY      326  GlnLysSerArgCysLeuArgAsnLeuMetLysThrProLeuPheValIleThrCys 345
DB      190  CAGGATCCAGGTGCTTGAGAAATCTGATGAGGACCCCTCTCTCTGCTGGTGTAACTGT 249

QY      346  AlaIleGlnMetGlyGluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThr 365
DB      250  GCCATCCAGATGGCAGTGGAGAAATCCAAAGCTCACACTCAAACCATGCTCTTCCAAACC 309

QY      366  PheTyrAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaIleSerAsp 385
DB      310  TTCTAGACCTCTGTATACAGAAAAACAGCGGCAGACACAGTGGAGAACTTCAGGTGAT 369

QY      386  PheIleArgSerLeuAspHisArgGlyAspLeuAlaLeuGluGlyValPheSerHisLys 405
DB      370  TTGTTCAGAGCTAGACTACTGTGGAGACCTGGCCCTGGAGGTGTGTCTCCACAG 429

QY      406  PheAspPheGluLeuGlnAspValSerSerValAsnGluAspValLeuLeuThrThrGly 425
DB      430  TTTGATTTCGAATTCAGCATGTGTGCAGCATGAATGAGGACGTCTCTGGTGGAGCGGG 489

QY      426  LeuLeuCysLysTyrThrAlaGlnArgPheLysProLysTyrLysPheHisLysSer 445
DB      490  CTCCTCTGTAAGTACACGCTCAGAGGCTGAGGCCCCACGCTATATAATCTTTATAATCC 549

QY      446  PheGlnGluTyrThrAlaGlyArgArgLeuSerSerLeuThrSerHisGluProGlu 465
DB      550  TTTCAGGAGTACACAGCAGCTCGGAGCTTCAGCATGTTTGTGTGAGTCCAGAGGCGAG 609

QY      466  GluValThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSer 485
DB      610  GAGGTGAGCAAGGGGAATAGTACTTGAAGAAAATGCTTCCATCTCCGACATCATCATCC 669

QY      486  ThrTyrSerSerLeuLeuArgTyrThr 494
DB      670  CTGTATGGTAATCTGCTCTGTACACT 696

```

RESULT 4

```

BB627584
LOCUS
DEFINITION
musculus CDNA clone 9530011P19 5', mRNA sequence.
ACCESSION
BB627584
VERSION
BB627584.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1..650
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="9530011P19"
/sex="male"
/tissue_type="urinary bladder"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male urinary
bladder"
/note="Site 1: Sall; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was

```

FEATURES
source

primed with a primer [5'
GAGAGAGAGAGATCCAGAGACTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 370.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAATAATTATCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from
Lambda FLC I."

BASE COUNT	164 a	174 c	173 g	139 t
ORIGIN				

Alignment Scores:	
Pred. No.:	2,086-83
Score:	836.00
Percent Similarity:	83.33%
Best Local Similarity:	75.93%
Query Match:	15.88%
DB:	10
Gaps:	0
Indels:	0
Mismatches:	36
Conservative:	164
Matches:	650
Length:	650

US-09-864-921-97 (1-1024) x BB627584 (1-650)

	Qy	129	AnLeuLysSerThrPheThrGluProIleLeuTrpArgLysAspGlnHisHisArg	148
	Db	2	AATCTGGAGAAACCTTCACAGAACCCTATCATGTGGAAGAGCAACCATCGTCAACCGT	61
	Qy	149	ValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleIleGlu	168
	Db	62	GTTGGAGCAGCTGACTTTGGGCAGCCCTGCTCGAGGCTCTGAAGAGCCCCCTGCCGTATTGAA	121
	Qy	169	GlyGluSerGlyLysGlyLysSerThrIleuLeuGlnArgIleAlaMetLeuTrpGlySer	188
	Db	122	GGCAGCTCTGGCAAAGGGAAGTCCACCCTGCTGCAGAGAATCGTATGCTCTGGGCCTCT	181
	Qy	189	GlyLysCysLysAlaLeuThrLysPheLysPheValPhePheLeuArgLeuSerArgAla	208
	Db	182	GGGGGCTGCAGGGCTCTGAAGGGGTTCAGATTAGTCTTCTTCAACCTCCAGAGCGCC	241
	Qy	209	GlnGlyGlyLeuPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThrIleArg	228
	Db	242	ATGGGGGACATTTCGAAACACTGTACGATCAGCTCTGGAACATACCCGACTTCATCAGC	301
	Qy	229	LysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgValLeuPheLeuAsp	248
	Db	302	AAGCCGACCTTCCAAGGCTCTGCTGTAAGCTACACAAGGAGGTCTCTTTCTTCTCGAT	361
	Qy	249	GlyTyraAsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysGluAsn	268
	Db	362	GTTTACATGAATTCATCCCAGAACCTGCCCGAATAATGAAGCCCTGATAAAGGANAAC	421
	Qy	269	HisArgPheLysAsnMetValIleValThrThrThrGluCysLeuArgHisIleArg	288
	Db	422	CATCGCTTCAAGAACATGGTCATGTCACCACCAACCGAGTGCTCGTAGGCATATCAGA	481
	Qy	289	GlnPheGlyAlaLeuThrAlaGluValGlyAspMetThrGluaspSerAlaGlnAlaLeu	308
	Db	482	CATGTTGGCGCCCTGACTCGCGAGGTGGGAGATATGCCGAAGACAGTGCCTCAAGATCTTC	541
	Qy	309	IleArgGluValLeuIleLysGluLeuAlaGluGlyLeuLeuGlnIleGlnLysSer	328
	Db	542	ATCGAGGCAGTCTGGTACTTCATGAGTGGTGAACGCTGTGGGCCCCCAATCAGGAGTCC	601
	Qy	329	ArgCysLeuArgAsnLeuMetLysThrProLeuPheValValIleThr	344
	Db	602	AGGTGCTGAGAAATCTCATGAAGACCCCTCTCTTCGTGTGTATCACC	649

RESULT 5				
BI824482				
LOCUS	BI824482	741 bp	mRNA	linear
DEFINITION	603038854F1 NIH MGC_115 Homo sapiens cDNA clone IMAGE:5179909 5', mRNA sequence.			EST 04-OCT-2001

ACCESSION	REFERENCE
VERSION	AUTHORS
KEYWORDS	TITLE
SOURCE	JOURNAL
ORGANISM	COMMENT

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM11448 row: k column: 14
 High quality sequence start: 3
 High quality sequence stop: 705

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

BASE COUNT	233 a	147 c	164 g	197 t
ORIGIN				
Alignment Scores:				
Pred. No.:	1.01e-80			
Score:	815.00			
Percent Similarity:	96.55%			
Best Local Similarity:	95.98%			
Query Match:	15.44%			
DB:	12			
	DB:			
	Gaps:			
	Matches:			
	Length:			
	Mismatch:			
	Conservat:			

US-09-864-921-97 (1-1024) x BI824482 (1-741)

Qy	1	MetAsn	Phel	Ile	Lys	Asp	Asn	Ser	Arg	--	Ala	Leu	Ile	Gln	Arg	Met	Gly	Met	Thr	Val	20
Db	228	ATG	AA	T	T	T	C	A	A	G	G	A	C	A	A	T	A	G	C	G	287
Qy	20	Ile	Lys	Gln	Ile	Thr	Asp	Asp	Leu	Phe	Val	Trp	Asn	Val	Leu	Asn	Arg	Glu	Glu	Val	40
Db	288	TAA	AG	C	NA	A	T	C	A	G	A	T	C	A	T	T	T	G	T	G	347
Qy	40	Ile	Lys	Gln	Ile	Thr	Asp	Asp	Leu	Phe	Val	Trp	Asn	Val	Leu	Asn	Arg	Glu	Glu	Val	60
Db	348	TCA	T	T	T	G	C	T	G	C	A	G	A	G	A	T	G	C	T	G	407
Qy	60	Lys	Gln	Ile	Thr	Asp	Asp	Leu	Phe	Val	Trp	Asn	Val	Leu	Asn	Arg	Glu	Glu	Val	80	
Db	408	AAA	AG	G	T	T	C	A	G	A	T	C	A	T	T	T	T	T	T	T	467
Qy	80	eup	Phc	Lys	Asp	Leu	Asn	Gly	Gln	Ser	Leu	Phe	Ile	Gln	Thr	Ser	Glu	Gly	Asp	Leu	100
Db	468	TAT	T	T	T	C	A	G	A	C	T	T	T	T	T	C	A	T	C	A	527

```

QY 100 spLeuAlaGlnAspLeu-LysAspLeuTyHisThrProSerPheLeuAenPheTyPro 119
Db 528 ATTGGCTCAGGATTACAGGACITGTACCATCCCATCTTCTGAACTTTATCC 587
QY 120 LeuGlyGluAspIleAspIlelePheAenLeuLysSerThrPheThrGluProIle-Le 139
Db 588 CTTGGTGAAGATATTGACATATTATTTTAACTTTGAAAGCACCTTCACAGAACCTGTGCTT 647
QY 139 utrPargLysAspGlnHisHieHieValGluGlnLeuThrLeuAen-GlyLeuLeuG 159
Db 648 GTGGAGGAAGGACCAACCATCAGCGCTGGAGCAGCTGACCTGAATGGTCTCCTGC 707
QY 159 InAlaLeuGlnSerProCysIleIleGluGly 169
Db 708 AGCTCTCCAGAG-CCTGCATCATTTGAAGG 738

RESULT 6
AV719179/c
LOCUS AV719179 GLC Homo sapiens cDNA clone GLCEQA10 5', mRNA sequence.
ACCESSION AV719179
VERSION AV719179.1 GI:10816331
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
AUTHORS Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
Y.,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA GLC clones
JOURNAL Unpublished
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..480
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCEQA10"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 138 a 120 c 85 g 137 t
ORIGIN

Alignment Scores:
Pred. No.: 1..89e-79 Length: 480
Score: 801.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.18% Indels: 0
DB: 9 Gaps: 0

US-09-864-921-97 (1-1024) x AV719179 (1-480)

QY 869 HisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAlaLeuMetLeuProTrp 888
Db 480 CATGAAGTATGACAGAGGATGAACGTGTAGAACAGCTCACCGCACTGATGTCGCTGG 421
QY 889 GlyCysAspValGlnGlySerLeuSerLeuLysHisLeuGluValProGln 908

```

```

Db 420 GGCTGTGACGTGCAAGGCAGCCTGAGCAGCCTGTGAAACATTTGGAGGAGGTCCACAA 361
QY 909 LeuValLysLeuGlyLeuLysAsnTrpArgLeuThrAspThrGluIleArgIleLeuGly 928
Db 360 CTCGTCAAGCTTGGGTGAAAACCTGGAGACTCACAGATACAGAGATTAGAAATTTAGGT 301
QY 929 AlaPhePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAenLeuAlaGlyAsnArg 948
Db 300 GCATTTTGGAAAGAACCCCTCTGAAAACCTTCAGCAGATTGAATTTGGCGGGAATCGT 241
QY 949 ValSerSerAspGlyTrpLeuAlaPheMetGlyValPheGluAenLeuLysGlnLeuVal 968
Db 240 GTGAGCAGTGTGATGGCTTGCCTTCATGGGTGATTGTGAGAACTCTTAAGCAATTAGTG 181
QY 969 PhePheAspPheSerThrLysGluPheLeuProAspProAlaLeuValArgLysLeuSer 988
Db 180 TTTTGTGACTTTAGTACTAAAGAATTTCTACTGATCCAGCATTAGTCAGAAAACCTTAGC 121
QY 989 GlnValLeuSerLysLeuThrPheLeuGlnGluAlaArgLeuValGlyTrpGlnPheAsp 1008
Db 120 CAAGTGTATCAAGTTAACTTTCTGCAAGAGCTAGGCTTGTGGGTGGCAATTTGAT 61
QY 1009 AspAspLeuSerValIleThrGlyAlaPheLysLeuValThrAla 1024
Db 60 GATGATGATCTCAGTGTATTACAGGAGCTTTTAACTAGTAACTGCT 13

RESULT 7
AV360053/c
LOCUS AV360053 602 bp DNA linear GSS 02-OCT-2000
DEFINITION IM0103H11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0103H11 F, genomic survey sequence.
ACCESSION AV360053
VERSION AV360053.1 GI:10473753
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
AUTHORS Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0103 row: H column: 11
Seq primer: CGTTGTAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 602.
FEATURES
source
1..602
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="UUGC1M0103H11"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource

```

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 139 a 153 c 139 g 171 t

ORIGIN

Alignment Scores:
Pred. No.: 4,598-79 Length: 602
Score: 799.00 Matches: 154
Percent Similarity: 88.21% Conservative: 18
Best Local Similarity: 78.97% Mismatches: 23
Query Match: 15.14% Indels: 0
DB: 28 Gaps: 0

US-09-864-921-97 (1-1024) x AZ360053 (1-602)

Qy 382 AlaAlaSerAspPheIleArgSerLeuAspHisArgGlyAspLeuAlaLeuGluGlyVal 401
Db 587 GCTTCAGGTGATTTTCCAGGAGCTTACTGCTGGAGACCTGGCCCTAGAGGTGTG 528
Qy 402 PheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAsnGluAspValLeu 421
Db 527 TTGCGCCCAAAATTTGATTTTGAACCCGAGCATGGTCCAGCATGAACGAGGACGCTG 468
Qy 422 LeuThrThrGlyLeuLeuCysLysThrAlaGlnArgPheLysProLysThrLysPhe 441
Db 467 GTGCAATAGGGCTCTCTGTAAGTACACAGCTCAGAGGCTGAAGCCACGATATAAATTC 408
Qy 442 PheHisLysSerPheGluThrAlaGlyArgArgLeuSerSerLeuThrSer 461
Db 407 TTTCATAAATCATTTTCAGGAGTACAGCGAGCTCGAGACTCAGCAGTTTCTGACGTC 348
Qy 462 HisGluProGluGluValThrLysGlyAsnGlyThrLeuGlnLysMetValSerLysSer 481
Db 347 AAAGAGCCAGAGGAGGTGAGCAAGGAAACAGCTACTTAAACAAATGGTTCCATCTCT 288
Qy 482 AspIleThrSerThrTyrosSerLeuLeuArgThrCysGlySerSerValGluAla 501
Db 287 GACATCATCCCTATATGGCAATCTGCTCTACAGTGTGGTCTGCACAGAGCA 228
Qy 502 ThrArgAlaValMetLysHisLeuAlaValThrGlnHisGlyCysLeuLeuGlyLeu 521
Db 227 ACCAGGGCGGTGATGAGGACCTTGCATGGTTTATCAGCAGGAGCTTACAGGACTT 168
Qy 522 SerIleAlaLysArgProLeuTrpArgGlnGlnSerLeuGlnSerValLysAsnThrThr 541
Db 167 TCAGTCACCAAGAGCGCTCTCTGGAGGAGGAATCAATCCAGAGTCTGAGAAATACCACT 108
Qy 542 GluGlnGluIleLeuLysAlaIleAsnLysLeuSerPheValGluCysGlyLysLeu 561
Db 107 GAGCAAGATGTTCTGAAGGCCATCAATGTAAATCTCTCGTAGAGTGTGGCATCAATTG 48
Qy 562 TyrGlnGluSerThrSerLysSerAlaLeuSerGlnGluPheGlu 576
Db 47 TTCTCAGAGAGTATGCTTAATCAGACCTGAGCCAGAAATTGAA 3

RESULT 8

BI908869

LOCUS

DEFINITION 603066455F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5215669 5',

mRNA sequence.

ACCESSION BI908869
VERSION BI908869.1 GI:16171950

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 748)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: csapbs@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LUAM1541 row: m column: 14

High quality sequence start: 7

High quality sequence stop: 744.

FEATURES

Location/Qualifiers

1..748

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5215669"

/tissue_type="leukocyte"

/lab_host="DH10B"

/clone_lib="NIH_MGC_118"

/notes="vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV

(destroyed); RNA source leukocytes from anonymous pool of

non-activated adult donors. Library is oligo-dT primed

and directionally cloned (EcoRV site is destroyed upon

cloning). Average insert size 1.7 kb, insert size range

1.2-3.3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 027. Note:

this is a NIH MGC Library."

BASE COUNT 236 a 148 c 166 g 198 t

ORIGIN

Alignment Scores:

Pred. No.: 1,97e-76 Length: 748

Score: 777.00 Matches: 159

Percent Similarity: 95.81% Conservative: 1

Best Local Similarity: 95.21% Mismatches: 5

Query Match: 14.72% Indels: 4

DB: 12 Gaps: 0

US-09-864-921-97 (1-1024) x BI908869 (1-748)

Qy 1 MethAsnPhelIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyMetThrValIle 20

Db 252 ATGAATTTCTATAAGGACAAATAGCCGAGCCCTTATTCAAAGAAATGGGAATGACTGTATA 311

Qy 21 LysGlnIleThrAspAspLeuPheValTrpAsnValLeuAsnArgGluGluValAsnIle 40

Db 312 AAGCAATATCAGATGACCTATTGTATGGATGTTCTGAATCCGAGGAAGTAACATC 371

Qy 41 IleCysCysGluLysValGluGlnAspAlaAlaArgGlyIleIleHisMetIleLeuLys 60

Db 372 ATTTGCTCGGAGAGGTGGAGCAGATGCTGCTAGAGGATCATTCACATGATTTGAAA 431

Qy 61 LysGlySerGluSerCysAsnLeuPheIleLysSerLeuLysGluTrpAsnTyrProLeu 80

Db 432 AAGGTTTCAGAGTCTCTTAACCTCTTTCTTAAATCCCTTAAGGAGTGAATCTCTCTTA 491

Qy 81 PheGlnAspLeuAsnGlyGlnSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100

Db 492 TTTCCAGGACTTGTAATGGCAAAAGCTCTTTTTCATCAGACATCAGAGGAGACTTGGACGAT 551
 QY 101 LeuAlaGluAspLeu--LysAspLeuTyHisThrProSerPheLeuAsnPheTyProL 120
 Db 552 TTGGCTCAGAGTTTCAAGAGGACTTGACCATACCCCATCTTTTCTGAACTTTTATCCCC 611
 QY 120 euGlyGluAspLeuAspLeuLeuPheAsnLeuLysSerThrPheThrGluProIleLeu 140
 Db 612 TTGGTCAAGATATTGACATATTATTTTAACCTTGAAGAGCCTTTCACAGAGCTGTCTGT 671
 QY 140 rpArgLysAspGlnHisHisArgValGluGlnLeuThrLeuAsnGlyLeuLeuGlnA 160
 Db 672 GGAGGAGGAGCAAAAC-CATCACCAGGCTGAGCAGGCTGACCTGAATGGCTC-CTGCAGG 729
 QY 160 laLeuGlnSerProCys 165
 Db 730 CTCTTCAGAGCCCTGTC 746

RESULT 9

BH348412 518 bp DNA linear GSS 03-DEC-2001
 LOCUS CH230-42F7.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
 DEFINITION CH230-42F7, genomic survey sequence.

ACCESSION BH348412
 VERSION 1
 KEYWORDS GSS.

SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 518)
 AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsengay,G., Geer,K., Shvartsbeyn
 ,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
 Jong,P. and Fraser,C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment
 JOURNAL Unpublished
 COMMENT Other GSSs: CH230-42F7.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/or ering information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 42 row: F column: 7
 Seq primer: T7
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1..518

/organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SsNHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-42F7"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 1"
 /notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
 CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
 Pieter de Jong"

BASE COUNT 133 a 124 c 146 g 115 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.2e-67 Length: 518
 Score: 693.00 Matches: 136

Percent Similarity: 86.71% Conservative: 14
 Best Local Similarity: 78.61% Mismatches: 22
 Query Match: 13.13% Indels: 1
 DB: 28 Gaps: 0

US-09-864-921-97 (1-1024) x BH348412 (1-518)

QY 261 lIeGluAlaLeuIleLysGluAsnHisArgPheLysAsnMetValIleValThrThr 280
 Db 2 ATCGAGGCCCTGGTAAGGAAACCATCGTTTCAGACATCGTTCATGTGCACCACC 61
 QY 281 ThrGluCysLeuArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
 Db 62 ACGGAGTGCCTCAGGCACATCAGACACAGCTGGCGCCTGACTGTGGAGGTGGAGATATG 121
 QY 301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGluLeuAlaGluGly 320
 Db 122 ACCGAAGACAGCGCCCGAGTTCATCCGGGAAGTGTCTGATAAATGAACCTGGCTGAAGGC 181
 QY 321 LeuLeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProLeuPhe 340
 Db 182 TTGTTGTTCCAGATCGAGGAGTCCAGGTGCTTGAGAAATCTGATGAGGACCCCTCTCTTC 241
 QY 341 ValValIleThrCys-AlaIleGlnMetGlySerGluPheHisSerHisThrGlnTh 360
 Db 242 GTGGTGATAACCTGTGACCATCCAGATGGCAGTGAGGAATTCGAAGCTCACACTCAAAC 301
 QY 360 rThrLeuPheHisThrPheTyAspLeuIleGlnLysAsnLysHisLysHisLysG 380
 Db 302 CATGCTCTTCCAAACCTTCTACGACCTCTCTGATACAGAAAAACAGCGCGACACAGTGG 361
 QY 380 yValAlaAlaSerAspPheIleArgSerLeuAspHisArgGlyAspLeuAlaLeuGluG 400
 Db 362 AGGAACCTCAGTGATTTTGTTCAGGAGCCTAGACTACTGTGGAGACCTGGCCCTGGAAGG 421
 QY 400 yValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAsnGluAspVa 420
 Db 422 TGTGTTCTCCCAAGTTTGATTTCGAACTTGAGGATGTGTCAGCATGATGAGGACGT 481
 QY 420 lleuLeuThrThrGlyLeuLeuCysLysTyThrAla 432
 Db 482 CCTGGTGAGGACGGGGCTCTCTGTGTAAGTACAGCG 518

RESULT 10

BH267158/6
 LOCUS

DEFINITION

ACCESSION BH267158

VERSION BH267158.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other GSSs: CH230-19B22.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources
(http://www.choi.org/bacpac/or ering information.html). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 19 row: B column: 22
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers

1..619
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-19B22"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SsHsd/MCW) BAC library produced by
Pieter de Jong"

BASE COUNT 163 a 141 c 158 g 157 t
ORIGIN

Alignment Scores:

Pred. No.: 3,12e-57 Length: 619
Score: 605.50 Matches: 119
Percent Similarity: 80.34% Conservative: 24
Best Local Similarity: 66.85% Mismatches: 30
Query Match: 11.47% Indels: 5
DB: 28 Gaps: 2

US-09-864-921-97 (1-1024) x BH267158 (1-619)

QY 73 LeuLysGluThrPAsnTyrProLeuPheGln-----AspLeuAsnGlyGlnSerLeu 89
DB 530 CTTGAGATGCAATCATCCCTCTATACATCATATATCTCTAACAGT-----CTT 477
QY 90 PheHisGlnThrSerGluGlyAspLeuAspLeuAlaGlnAspLeuLysAspLeuTyr 109
DB 476 TTTCATCAGAACTTGAAAGAACTTGATGTTCTGGCCAGAGTCTAAAGACTTATAC 417
QY 110 HisThrProSerPheLeuAsnPheTyrProLeuGlyGluAspLeuAspLeuPheAsn 129
DB 416 AACAGCCCTGTTTAAAGAACTTCTTCTCTGGTGAGGATATCGACATCATTTTCAAT 357
QY 130 LeuLysSerThrPheThrGluProLeuLeuTyrArgLysAspGlnHisHisArgVal 149
DB 356 CTGCAGATCACCTTCACAGAACCTGTCTTGTGGAGGAGGACCATCGTCATCCGAGTA 297
QY 150 GluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleIleGluGly 169
DB 296 GAGCAGATGACCTTGGGAGGCTGTGGAGGCTCTGAAGAGTCCCTGCTCATTAAGGG 237
QY 170 GluSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAlaMetLeuTyrGlySerGly 189
DB 236 GAGTCTGGCAAGGGAAGTCCACCTCTGTCGCAAAAAATTGCCATGCTCTGGCCCTATGA 177
QY 190 LysCysLysAlaLeuThrLysPheLysPheValPhePheLeuArgLeuSerArgAlaGln 209
DB 176 ATGTGCCAGCTCTGAACAGTTCAAATTTGGTCTTTTCATCCGCTGAGCAGTGTAGG 117
QY 210 GlyGlyLeuPheGluThrLeuCysAspGlnLeuLeuAspLeuProGlyThrIleArgLys 229
DB 116 GTTGGCTCTTTGAACATGATGATCATAGCTCGTGAACATACCTGACTCCATCAGCAA 57
QY 230 GlnThrPheMetAlaMetLeuLysLeuArgGlnArgValLeuPheLeuLeu 247
DB 56 CCAACCTTCAGGCTCTGCTGCTGAAGTTACACAAGAAAGTCTCTTCTCTCTC 3

RESULT 11

AV656315 371 bp mRNA linear EST 16-JAN-2002
LOCUS
DEFINITION AV656315 GLC Homo sapiens cDNA clone GLCEQA10 3', mRNA sequence.
ACCESSION AV656315

AV656315.1 GI:9877329

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 371)

AUTHORS

Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,W., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.

Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

21625106

JOURNAL

MEDLINE 11752456

PUBMED

CONTACT: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

1..371 Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="GLCEQA10"

/tissue_type="corresponding non cancerous liver tissue"

/dev_stages="Adult"

/lab_host="SOLR"

/clone_lib="GLC"

/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 112 a 85 c 91 g 82 t 1 others

ORIGIN

Alignment Scores:
Pred. No.: 7.38e-57 Length: 371
Score: 599.00 Matches: 123
Percent Similarity: 87.23% Conservative: 0
Best Local Similarity: 87.23% Mismatches: 0
Query Match: 11.35% Indels: 18
DB: 9 Gaps: 1

US-09-864-921-97 (1-1024) x AV656315 (1-371)

QY 625 GluLysAlaAlaGluAspThrGlyGlyIleHisMetGluGluAlaProGluThrTyrIle 644
DB 1 GAAAGGCTGCAGAGACACAGGTGGAATCCACATGGAAGAGGCCGCCAGAACCTACATT 60
QY 645 ProSerArgAlaValSerLeuPheAsnTrpLysGlnGluPheArgThrLeuGluVal 664
DB 61 CCCAGAGGGCTGATCTTTCTTCACTGGAAGCAGGAATTCAGACTCTGGAGTTC 120
QY 665 ThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleArgTyrLeuGlyLysIlePhe 684
DB 121 ACACCTCCGGGATTTTCAGCAAGTTGAATAAGCAAGATATCAGATATCTGGGAAAAATATTC 180
QY 695 SerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyValAlaGlySerLeu 704
DB 181 AGCTCTGCCAAGCCTCAGGCTCAATAAAGAGATGTGCTGTGTGGTGTGGAGGCCTC 240
QY 705 SerLeuValLeuSerThrCysLysAsnIleTyrSerLeuMetValGluAlaSerProLeu 724
DB 241 AGTTTGGTCTTCAGCACTGTAAGAACATTTATTTCTCTCATGTGGAGCAGCCAGCTCC 300
QY 725 ThrIleGluAspGluArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis 744
DB 301 ACCATAGAAGATGAGAGGCACATCATCTCTGTAAACACCTG----- 342

```

QY 745 AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsn 764
Db 343 -----GACAGCTTGGTAACTTGAAGAAC 366

QY 765 Leu 765
Db 367 CTT 369

RESULT 12
AI263294/c 364 bp mRNA linear EST 03-FEB-1999
LOCUS qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
DEFINITION mRNA sequence.
ACCESSION AI263294
VERSION AI263294.1 GI:3871497
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 2146 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 364.
FEATURES
Location/Qualifiers
1..364
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2005417"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pan1"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
BASE COUNT 117 a 84 c 55 g 108 t
ORIGIN
Alignment Scores:
Pred. No.: 2,71e-55 Length: 364
Score: 585.00 Matches: 114
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.09% Indels: 0
DB: 9 Gaps: 0

US-09-864-921-97 (1-1024) x AI263294 (1-364)

QY 911 LysLeuGlyLeuLysAsnTrpArgLeuThrAspThrGluLeuArgIleLeuGlyAlaPhe 930
Db 364 AAGCTTGGGTGGAAGAACTGGAGACTCACAGATTACAGATTAGTGGCAATTT 305

QY 931 PheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnAlaGlyAsnArgValSer 950
Db 304 TTGGGAAGAACCTCTGAAAACTTCCAGCAGTTGAATTTGGCGGAATCTGTGAGC 245

QY 951 SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLeuValPhePhe 970
Db 244 AGTGATGGATGGCTTGCCTTCATGGGTGATTGAGAACTCTTAAGCAATTAGTGTGTTT 185

```

```

QY 971 AspPheSerThrLysGluPheLeuProAspProAlaLeuValArgLysLeuSerGlnVal 990
Db 184 GACTTTAGTACTAAGAATTTCTACTGATCCAGCATTAGTCAGAAACTTAGCCAAATG 125

QY 991 LeuSerLysLeuThrPheLeuGlnGlnAlaArgLeuValGlyTrpGlnPheAspAsp 1010
Db 124 TTATCCAGTTAACTTTTCTGCAAGAAGCTAGGCTTGTGGTGGCAATTTCATGATGAT 65

QY 1011 AspLeuSerValIleThrGlyAlaPheLysLeuValThrAla 1024
Db 64 GATCTCAGTGTATTACAGGTCGCTTTTAAACTAGTAACGTCT 23

RESULT 13
BI854236
LOCUS BI854236
DEFINITION 603381263F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5389239 5',
ACCESSION BI854236
VERSION BI854236.1 GI:15994983
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLA411991 row: m column: 16
High quality sequence stop: 764.
FEATURES
Location/Qualifiers
1..775
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:5389239"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam4"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furth,
NIH Reference for transgenic model: Li et al., Cell Growth
and Differentiation 7, 3-11 (1996)."
BASE COUNT 226 a 162 c 196 g 191 t
ORIGIN
Alignment Scores:
Pred. No.: 9.18e-55 Length: 775
Score: 585.00 Matches: 118
Percent Similarity: 78.66% Conservative: 11
Best Local Similarity: 71.95% Mismatches: 35
Query Match: 11.09% Indels: 0
DB: 12 Gaps: 0

US-09-864-921-97 (1-1024) x BI854236 (1-775)

QY 861 GluLysAspGlyAsnGluAlaLeuHisGluLeuIleAspArgMetAsnValLeuGln 880
Db 861 GATCTCAGTGTATTACAGGTCGCTTTTAAACTAGTAACGTCT 23

```

Db 2 GAAAGGATGGGAATGAAGCTCTACAGGAACATGATCGGCAGGCTTGGCGTCTCTGGAGAG 61

Qy 881 LeuThrAlaLeuMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900

Db 62 CTCACATACATGATGCTGCTGGTCTGGGATGTGCACACGAGCCTGCCAAGCTGTTG 121

Qy 901 LysHisLeuGluGluValProGlnLeuValLeuValLeuGlyLeuLysAsnTrpArgLeuThr 920

Db 122 AGCAGTTGGAGGGGACCCAGGACTTGCACAACTGGATTGAAAACCTGGAGACTCAGA 181

Qy 921 AspThrGluLeuArgLeuGlyAlaPhePheGlyLysAsnProLeuLysAsnPheGln 940

Db 182 GACGAAGAGATTAAAGATTAGTGAAATTTCTGGAGATGAATCTCTGAGAGACTTGCAG 241

Qy 941 GlnLeuAsnLeuAlaGlyAsnArgValSerSerAspGlyTrpLeuAlaPheMetGlyVal 960

Db 242 CAGTTGGATTAGCGGGGACCTGTGTGACAGTGCAGGATGGCTTTACTTTCATGAATGTG 301

Qy 961 PheGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPheLeuProAsp 980

Db 302 TTTGAGAACTGAGCAGATTAGTGTGTTTTTGACTTTAGCACTGAGGAGTTCTTACCGAT 361

Qy 981 ProAlaLeuValArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000

Db 362 GCAGCACTGCTGAGGAACCTTAGTCAAGTGTATCCAGTTAACTCTTCTGCAAGAGTGA 421

Qy 1001 ArgLeuValGlyTrpGlnPheAspAspAspLeuSerValIleThrGlyAlaPheLys 1020

Db 422 AAGCTCACGGCTGGGAGTTGATGACTATGATATTAGCGCTATTAAAGGCACCTTTAAA 481

Qy 1021 LeuValThrAla 1024

Db 482 CTAGTCACTGCT 493

RESULT 14

BH358172/c
LOCUS
DEFINITION
CH230-18E7.TJB CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-18E7, genomic survey sequence.

ACCESSION

VERSION
BH358172.1 GI:17288906

KEYWORDS

SOURCE
Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 817)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartabeyn
A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
Jong, P. and Fraser, C.M.

TITLE

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

JOURNAL

COMMENT

Other GSSs: CH230-18E7.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.choi.org).
Clones may be purchased from BACPAC Resources
(<http://www.chori.org/bacpac/orxing/information.html>). BAC end
page: http://www.tigr.org/tadb/bac_ends/rat/bac_end_intro.html
Plate: 18 row: E Column: 7

Seq primer: SP6

Class: BAC ends.

FEATURES

source
Location/Qualifiers

1..817

/organism="Rattus norvegicus"

/mol_type="genomic DNA"
/strain="BN/SENHed/MCW"
/db_xref="taxon:10116"
/clone="CH230-18E7"
/sex="Female"
/cell_type="Brain"

/clone_lib="CHORI-230 Segment 1"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SENHed/MCW) BAC library produced by
Pieter de Jong"

BASE COUNT 211 a 182 c 199 g 225 t

ORIGIN

Alignment Scores:
Pred. No.: 2,05e-52 Length: 817
Score: 564.50 Matches: 121
Percent Similarity: 74.6% Conservative: 25
Best Local Similarity: 64.36% Mismatches: 37
Query Match: 10.70% Indels: 7
DB: Gaps: 2

US-09-864-921-97 (1-1024) x BH358172 (1-817)

Qy 62 GlySerGluSerCysAsnLeuPheLeuLysSerLeuLysGluTrpAsnTy-ProLeuPhe 81

Db 556 GGCATCTCTGCCATTCATCTCCCAAGCCTTGGGCTTCGAGATGCAATCATGCTCTAT 497

Qy 82 Gln-----AspLeuAsnGlyGlnSerLeuPheHisGlnThrSerGluGlyAspLeu 98

Db 496 AACATCATATTTCTCTAACAGGT-----CTTTTTCATCAGAACTTGGAGAACTTG 443

Qy 99 AspAspLeuAlaGlnAspLeuLysAspLeuTyHisThrProSerPheLeuAsnPheTy 118

Db 442 GATGTTCTGGGCCAGAGTCTAAAGACTTATACACAGCCCTGTTTAAAGAACTTCITT 383

Qy 119 ProLeuGlyGluAspIleAspIlePheAsnLeuLysSerThrPheThrGluProIle 138

Db 382 CTTCTGGGTGAGGATATCGACATCATTTTCAATCTGCAGATCACCTTTCACAGAACCTGTC 323

Qy 139 LeuTrpArgLysAspGlnHisHisArgValGluGlnLeuThrLeuAsnGlyLeuLeu 158

Db 322 TTGTGGAGGAGGACCATCGTCATCACCGAGTAGAGCAGATGACCTGGCAGCCTGCTG 263

Qy 159 GlnAlaLeuGlnSerProCysIleIleGluGlyGluSerGlyLysGlyLysSerThrLeu 178

Db 262 GAGGCTCTGAAGAGTCCCTGCTCATTTGAAGGGAGTCTGGCAAGGGAAGTCCACCTG 203

Qy 179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeuThrLysPheLys 198

Db 202 CTGCAAAAATTCATCTCTGGGCTCTGGGATGTCGCCAGCTCTGAACCAAGTTCAA 143

Qy 199 PheValPhePheLeuArgLeuSerArgAlaGlnGlyLeuPheGluThrLeuCysAsp 218

Db 142 TTGTTCTTCTTCATCCGCTGAGCAGTCTAGGGGTGGCTGTTTGAACAACATTGTA-GAT 84

Qy 219 GlnLeuAspIleProGlyThrIleArgLysGlnThrPheMetAlaMetLeuLys 238

Db 83 CAGCTCGTGAACATACCTGACTCCATCAGCAAAACCACTTCAGGGCTCTGCTGCT-AAG 25

Qy 239 LeuArgGlnArgValLeuPheLeu 246

Db 24 TTACACAGAAGTCTCTTCTC 1

RESULT 15

BX109471/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BX109471
IMAGE:1909558, mRNA sequence.
BX109471
BX109471.1 GI:27835805
Homo sapiens (human)

503 bp
mRNA
linear
EST 07-FEB-2003
NCI CGAP Lu5 Homo sapiens cDNA clone IMAGE998B234682 ;
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 503)
 AUTHORS Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M., Radelof, U., Schneider, D. and Korn, B.
 Human Unigeneset - RZPD3
 Unpublished
 TITLE bin/showLib.pl.cgi/response?libNo=972
 COMMENT RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
 Contact: Ina Rolfes
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
 Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
 RZPD; IMAGp998B234682.
 RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
 Human Unigeneset - RZPD3 (RZPDLIB No.972)
 http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=972
 Contact: Ina Rolfes
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
 Heubnerweg 6, D-14059 Berlin, Germany
 Tel: +49 30 32639 101
 Fax: +49 30 32639 111
 www.rzpd.de

This clone is available royalty-free from RZPD;
 contact RZPD (clone@rzpd.de) for further information. Seq primer:
 M13r, Primer sequence: TTTCACAGGAACAGCTATGAC.

FEATURES

Location/Qualifiers
 1..503
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGp998B234682 ; IMAGE:1909558"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Lu5"
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 127 a 111 c 128 g 128 t 9 others
 ORIGIN

Alignment Scores:

Pred. No.: 1.38e-52 Length: 503
 Score: 563.00 Matches: 106
 Percent Similarity: 96.40% Conservative: 1
 Best Local Similarity: 95.50% Mismatches: 4
 Query Match: 10.67% Indels: 0
 DB: 13 Gaps: 0

US-09-864-921-97 (1-1024) x EX109471 (1-503)

QY 89 LeuPheHisGlnThrSerGluGlyAspLeuAspLeuAlaGlnAspLeuLysAspLeu 108
 Db 334 CTTTTCATCAGACATCAGAGGAGNCTTGGACGATTTCGCTCAGGATTTAAGGNCITG 275
 QY 109 TyrHisThrProSerPheLeuLeuPheTyrProLeuGlyGluAspIleAspIlelePhe 128
 Db 274 TNCATNCCCATCTTTTCTGAACCTTTATCCCTTGGTGAAGATATTGACATATTATTT 215
 QY 129 AsnLeuLysSerThrPheThrGluProIleLeuTrpArgLysAspGlnHisHisArg 148
 Db 214 AACCTTGAAGAAGCACCTTCACAGAACCTGCTCTGGAGGAGGACCAACACCATCACC 155
 QY 149 ValGluGlnLeuThrLeuLeuGlyLeuLeuGlnAlaLeuGlnSerProCysIleIleGlu 168
 Db 154 GTGGACAGCTGACCTTGATGGCTCTCTGCAGGCTCTTCAGAGCCCTCGATCATTTGA 95
 QY 169 GlyGluSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAlaMetLeuTrpGlySer 188
 Db 94 GGGGAATCTGGCAAGGCAAGTCCACTCTGCTGCAGGAATTGCCATGCTCTGGGGCTCC 35

QY 189 GlyLysCysLysAlaLeuThrLysPheLysPhe 199
 Db 34 GGAAAGTCAAGGCTCTGACCAAGTTCAAATTC 2

Search completed: December 13, 2003, 04:24:54
 Job time : 3346 secs

THIS PAGE BLANK (USPTO)